PCT/ SE 00 / 0 0 1 5 8

PRV

PATENT- OCH REGISTRERINGSVERKET
Patentavdelningen

REC'D 08 MAR 2000
WIPO

4

Intyg
Certificate

Härmed intygas att bifogade kopior överensstämmer med de handlingar som ursprungligen ingivits till Patent- och registreringsverket i nedannämnda ansökan.

This is to certify that the annexed is a true copy of the documents as originally filed with the Patent- and Registration Office in connection with the following patent application.

(71) Sökande — Lars Liljeryd, Solna SE
Applicant (s)

(21) Patentansökningsnummer — 9903552-9
Patent application number

(86) Ingivningsdatum — 1999-10-01
Date of filing

(30) Prioritet begärd från — 1999-01-27 SE 9900256-0

Stockholm, 2000-03-02

För Patent- och registreringsverket
For the Patent- and Registration Office

A. Södervall
Anita Södervall

Avgift
Fee

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

PATENT- OCH REGISTRERINGSVERKET
SWEDEN

Postadress/Adress: Box 5055, S-102 42 STOCKHOLM
Telefon/Phone: +46 8 782 25 00, Vx 08-782 25 00
Telex: 17978 PATOREG S
Telefax: +46 8 666 02 86, 08-666 02 86

EFFICIENT SPECTRAL ENVELOPE CODING USING DYNAMIC SCALEFACTOR GROUPING AND TIME/FREQUENCY SWITCHING

TECHNICAL FIELD

The present invention relates to a new method and apparatus for efficient coding of spectral envelopes or scalefactors in audio coding systems. The method may be used both for natural audio coding and speech coding and is especially suited for coders using SBR [WO 98/57436].

BACKGROUND OF THE INVENTION

Audio source coding techniques can be divided into two classes: natural audio coding and speech coding. Natural audio coding is commonly used for music or arbitrary signals at medium bitrates, and generally offers wide audio bandwidth. Speech coders are basically limited to speech reproduction but can on the other hand be used at very low bitrates, albeit with low audio bandwidth. In both classes, the signal is generally separated into two major signal components, the "spectral envelope" representation and the corresponding "residual" signal. Throughout the following description, the term "residual" refers to the error signal obtained from linear predictive coding (LPC) as well as normalised sample values obtained from filter banks or time-to-frequency transforms, whereas the term "spectral envelope" refers to a set of values obtained on a possibly larger timeframe, for example the filter coefficients from the LPC or the values used for normalisation of the residual. At medium bitrates, the residual constitutes the main part of the bitstream while the spectral envelope part is merely a fraction. At very low bitrates, the spectral envelope constitutes a comparably larger part of the bitstream. Hence, it is indeed important to represent the spectral envelope compactly when using lower bitrates.

In audio coders, the spectral envelope representation is segmented into granules. Prior art systems use static, relatively short, granule-lengths to achieve good temporal resolution. However, this prevents from optimal utilisation of the frequency domain masking known from pshycho-acoustics, thus limiting the coding gain. To improve coding gain, and still achieve good temporal resolution during transient passages, modern coders employs adaptive window switching, i.e. they switch granule-lengths depending on the signals statistics. Unfortunately, long transition windows, with poor frequency selectivity, are needed to transform the granule lengths. These windows further reduce the overall coding gain, hence minimum usage of the short granule-lengths is a prerequisite for maximum coding gain.

The spectral envelope is a function of two variables: time and frequency. The encoding can be done by exploiting redundancy in either direction of the time/frequency plane. Generally, coding of the spectral envelope is performed in the frequency direction using delta coding (DPCM), linear prediction (LPC), or vector quantization (VQ).

SUMMARY OF THE INVENTION

The present invention provides a new method and an apparatus for spectral envelope encoding. The invention teaches how to perform and compactly signal a time/frequency mapping of the envelope representation, and further,

encode the spectral envelope data efficiently using adaptive time/frequency direction coding. The invention exploits the fact that adjacent transient passages of the signal, needing high temporal resolution coding, are separated at least by a minimum time, T_{nmin}. In the encoder, a transient detector decides whether the current granule contains transients, and if so, determines the position of the onset of the transient on a relatively short time basis. This position, if any, is encoded and sent to the decoder. Both the encoder and decoder share rules that specify the time/frequency distribution of the spectral envelope samples, given a certain combination of subsequent transient positions. The rules can be realised as a book of tables explicitly specifying the division of the current granule in terms of samples in the time/frequency plane. In the absence of transients, i.e. for quasistationary signals, a time/frequency grid with low temporal and high frequency resolution is used as default. In the vicinity of transients, the temporal resolution is increased at the expense of frequency resolution.

The method is also applicable on envelope encoding based on prediction. Instead of grouping subband samples, predictor coefficients are generated for the segments according to the indexing system. Different predictor orders may be used for transient and quasi-stationary (tonal) segments.

The present invention presents a new and efficient method for scalefactor redundancy coding. A dirac pulse in the time domain transforms to a constant in the frequency domain, and a dirac in the frequency domain, i.e. a single sinusoid, corresponds to a signal with constant magnitude in the time domain. Simplified, on a short term basis, the signal shows less variations in one domain than the other. Hence, using prediction or delta coding, coding efficiency is increased if the spectral envelope is coded in either time- or frequency-direction depending on the signal characteristics.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention will now be described by way of illustrative examples, not limiting the scope or spirit of the invention, with reference to the accompanying drawings, in which:

Fig. 1a illustrates uniform sampling in time of the spectral envelope.
Fig. 1b illustrates non- uniform sampling in time of the spectral envelope.
Fig. 2 illustrates possible transient positions in granules of two sizes, assuming a minimum time between consecutive transients.
Fig. 3 illustrates transient detector lookahead and granule interdependency.
Fig. 4 are some examples of subgranules grouped into time segments.
Fig. 5 illustrates time-/frequency switched envelope coding.
Fig. 6 shows a block diagram of an encoder using the envelope coding according to the invention.
Fig. 7 shows a block diagram of a decoder using the envelope coding according to the invention.

DESCRIPTION OF PREFERRED EMBODIMENTS

Scalefactor Generation Scheme

In conventional subband coders, the subband samples obtained from the analysis filterbank are converted into scalefactors and scaled subband samples in the encoder. There are many approaches to this conversion and the scalefactors may or may not represent the spectral envelope of the signal and the terms are to be interpreted in a general way. However, most systems, except for those employing PNS ["Improving Audio Codecs by Noise Substitution", D. Schultz, JAES, vol. 44, no. 7/8, 1996], have in common that both scalefactors and scaled subbandsamples are transmitted and combined during the synthesis at the decoder. With SBR this is not the case (considering the highband), only the spectral course structure needs to be transmitted, which in some coders corresponds to a transmittal of the scalefactors only. This puts higher demands on how to generate scalefactors, since the scaled subband samples, in particular their temporal information, no longer is available. The problem will now be demonstrated by means of an example:

Fig 1 shows the time-/frequency representation of a musical signal where sustained chords are combined with sharp transients with mainly high frequency contents. In the lowband the chords have high energy and the transient energy is low, whereas the opposite is true in the highband. The scalefactors that are generated during time intervals where transients are present are dominated by the high intermittent transient energy. At the SBR process in the decoder, the spectral envelope of the transposed signal is estimated using the same instantaneous time-/frequency resolution that was used for the analysis of the original highband. The amplification factors in the envelope adjusting filterbank are calculated as the quotients between the scalefactors. For this kind of signal, a problem arises: The transposed signal has the same ratio between chord and transient energies as the lowband. The gains needed in order to adjust the transposed transients to the correct level thus cause the transposed chords to be amplified relative the original highband level for the full duration of the scalefactor containing transient energy. These momentarily too loud chord fragments are perceived as pre- and post echoes to the transient, see Fig 1a. This kind of distortion will be referred to as gain induced pre- and post echoes. The phenomenon can be eliminated by constantly updating the scalefactors at such a high rate that the time between an update and an arbitrarily located transient is guaranteed to be short enough not to be resolved by the human hearing. However, this approach would drastically increase the amount of data to be transmitted and is thus not practical.

Therefore a new system for scalefactor generation is presented. The principal solution is to maintain a low update rate during tonal passages, which make up the majority of a typical programme material, and by means of a transient detector localize the transient positions, and update the scalefactors close to the leading flanks, see Fig 1b. This eliminates gain induced pre-echoes. In order to well represent the decay of the transients, the update rate is momentarily increased in a time interval after the transient start. This eliminates gain induced post-echoes. The time grouping during the decay is not as crucial as finding the start of the transient, as will be explained below. In order to compensate for the smaller time steps, larger groups in frequency are used during the transient, keeping the data size within limits. Moreover, it is possible to use an analysis by synthesis approach, i.e. having a decoder in the encoder to assess the most beneficial time/frequency sampling.

Notice that in this example the varying time and frequency sampling is obtained by grouping of the subband samples from a fixed filterbank in different ways. Variable time-/frequency resolution is employed by some conventional subband coders as well. The difference lies partly in the switching criteria, partly in that they in general also switch the filterbank size. Such a change in size can not take place immediately, so called transition windows are needed, and thus the update points can not be chosen as freely as when the filterbank remains unchanged. Furthermore, in the SBR case a high frequency resolution of the envelope for tonal signals is desired all the way up to the start of the transients. This demand is met by keeping the filterbank size constant, since at every moment the high resolution subband samples are available. Obviously, for traditional coding, this type of design imposes a compromise between time and frequency resolution of the filterbank. However, when using SBR, information from the bank is always discarded, hence, the filterbank can be designed to meet both the highest temporal and frequency resolution needed.

In order to correctly interpret the received envelope data, the selected scalefactor grouping must be signalled. Typical coders operate on a block basis, where every block represents a fixed time interval. Those blocks will be referred to as granules. If a non-uniform sampling according to Fig 1c is to be employed, the problem of scalefactor segments spanning over the granule borders must be dealt with. Furthermore, the signalling must be flexible enough to cover all combinations of interest, without generating a too large amount of control data.

Assume that granule has a length of q time quantization steps, hereinafter called subgranules. Theoretically, transients can occur in C combinations, ranging from no transient at all to q transients, where C is given by

$$C=\sum_{n=0}^{q}\binom{q}{n}=2^{q} \qquad \text{(Eq 1)}$$

In order to signal C states, $\ln2(C) = \ln2(2^q) = q$ bits are required, corresponding to one bit per quantization step within the granule. If different frequency groupings is to be used in the segments, even more bits might be required in order to signal the frequency resolution chosen. In low bitrate applications the number of control signal bits must be kept at a minimum. As will be shown below, many of the above states are not very likely, and would also correspond to too large amounts of scalefactors to be practical at the limited bitrate. According to the present invention, the number of states to be signalled can be reduced significantly with little or no sacrifice of quality for practical signals if the following simplifications are made:

1. Only the transient start position needs to be transmitted. The time and frequency grouping around this position can be dealt with by employing a set of rules in the encoder and decoder, which are based on the properties of typical transients.
2. There exists a fixed minimum time between consecutive transients, i.e. transients can not be arbitrary close to each other. It is thus possible to introduce a blocking time in the transient detection/signalling system, reducing the number of states.

The minimum time between consecutive transients in music programme material can be estimated in this way: In musical notation, the rhythmic "pulse" is described by a time signature expressed as a fraction A/B, where A denotes the number of "beats" per bar and $1/B$ is the type of note corresponding to one beat, for example a ¼ note,

commonly referred to as a quarter note. Let t denote the tempo in Beats Per Minute (BPM). The time per note of type $1/C$ is then given by

$$T_n = (60/t)*(B/C)\ [s] \qquad \text{(Eq 2)}$$

Most music pieces fall within the 70 – 160 BPM range, and in 4/4 time signature the fastest rhythmical patterns are for most practical cases made up from 1/32 or 32:nd notes. This yields a minimum time $T_{nmin} = (60/160)*(4/32) = 47$ ms. Of course lower time periods than this may occur, but such fast sequences (>21 tones per second) almost get the character of buzz and need not be fully resolved.

The necessary time resolution T_q must also be established. In some cases a transient original signal has its main energy in the SBR highband. This means that the encoded spectral envelope must carry all the "timing" information. The desired timing precision thus determines the resolution needed for encoding of leading flanks. T_q is much smaller than the minimum note period T_{nmin}, since small time deviations within the period clearly can be heard. In most cases however, the transient has significant energy in the lowband. The above described gain-induced pre-echoes must fall within the so called pre- or backward masking time T_m of the human auditory system in order to be inaudible. Hence T_q must satisfy two conditions:

$$T_q << T_{nmin} \qquad \text{(Eq 3)}$$

$$T_q < T_m \qquad \text{(Eq 4)}$$

Obviously $T_m < T_{nmin}$ (otherwise the notes would be so fast that they could not be resolved) and according to ["Modeling the Additivity of Nonsimultaneous Masking", Hearing Res., vol. 80, pp. 105-118 (1994)], T_m amounts to 10-20 ms. Since T_{nmin} is in the 50ms range, a reasonable selection of T_q according to Eq 3 results in that the second condition is also met. Of course the precision of the transient detection in the encoder and the time resolution of the analysis/synthesis filterbank must also be considered when selecting T_q.

Tracking of trailing flanks is less crucial, for several reasons: First, the note-off position has little or no effect on the perceived rhythm. Second, most instruments do not exhibit sharp trailing flanks, but rather a smooth decay curve, i.e. a well defined not-off time does not exist. Third, the post- or forward masking time is substantially longer than the pre masking time.

Given the above established design limits, the selection of granule length presents a problem in itself. In Fig 2 granule lengths of $8T_q$ and $16T_q$, where $8T_q <= T_{nmin}$, are compared. The subgranule number is shown in the top two rows, and T denotes that a transient is present within the subgranule. Since input signal tones are separated at least Tnmin, the maximum number of transients within a granule is 1 and 2 respectively. The number of possible signal combinations during the time $16T_q$ is 53 and the combinations seen by one granule can be grouped as follows:

	Granule length $8T_q$:	Granule length $16T_q$:
0 transient:	1 case	1 case
1 transient:	8 cases	16 cases
2 transients:	0 cases	36 cases (not 16 over 2, due to T_{nmin})

If a transient is present within a granule, the position(s) must be signalled. Assuming a system where this control signal is only sent in case of a transient, the required number of control bits is respectively

$$B_{1dyn} = \text{ceil}\{\ln2(\text{number of transient cases})\} = \text{ceil}\{\ln2(8)\} = \text{ceil}\{3\} = 3, \text{ and} \qquad \text{(Eq 5)}$$

$$B_{2dyn} = \text{ceil}\{\ln2(\text{number of transient cases})\} = \text{ceil}\{\ln2(16 + 36)\} = \text{ceil}\{5.7\} = 6, \qquad \text{(Eq 6)}$$

where ceil{·} denotes round to nearest higher integer.

In order to calculate the average control signal bitrate, the likelihood of the different cases is needed, and this is in general unknown. However, note periods near the limit T_{nmin} are not very common and for the cases where only one transient is present within the time frame $16T_q$, the shorter granule system is cheaper by a factor 2, since no control signal is transmitted for half the number of granules. On the other hand, in a system where the number of control signal bits must be fixed, the bit demands are

$$B_{1fix} = \text{ceil}\{\ln2(\text{total number of cases})\} = \text{ceil}\{\ln2(1 + 8)\} = \text{ceil}\{3.2\} = 4, \text{ and} \qquad \text{(Eq 7)}$$

$$B_{2fix} = \text{ceil}\{\ln2(\text{total number of cases})\} = \text{ceil}\{\ln2(1 + 16 + 36)\} = \text{ceil}\{5.7\} = 6. \qquad \text{(Eq 8)}$$

In this case the $16T_q$ system is a factor 6/4 = 1.5 cheaper. (Note that signalling of 9 states with 4 bits is not ideal. The granule could instead be divided into 7 steps, which would require 3 bits, or the remaining 7 states of the 4 bit signal could be used for other purposes, as shown below.) To put this into perspective, a hypothetic low bitrate envelope encoder is studied: Assume granules of length $16T_q <= T_{nmin}$, which is a more practical selection than in the above examples, and that the control signal always is sent. The signalling costs are $B_{fix} = \text{ceil}\{\ln2(1+16)\} = 5$ and $B_{gen} = q = 16$ for the totally flexible system. The saving by using the transient indexing system is 16 – 5 = 11 bits. A typical average number of scalefactors per frame is 40 and the average number of bits per scalefactor is 3 (due to lossless coding). The saving is thus about 4 scalefactors corresponding to 10 % of the envelope data, i.e. it is significant at such low bitrates.

According to the present invention, the above transient start information is used for implicit signalling of segment borders and frequency resolutions immediately after/between transients. This will now be described, using the above $8T_q$ system with fixed number of control bits as an example. The four control bits are divided into two signals; three bits are used to signal the transient position, *tran_pos*, and one bit to signal the presence of a transient, *tran_flag*. These values are used in combination with transient position values from preceding granules to determine the time/frequency grid to be used for the current granule. These grids are stored in tables that are available to both the encoder and decoder. Given the common tables and the signalling of the transient positions ensures unambiguous decoding of the envelope energy values. In applications where there are non-critical delay restrictions, as in point to multipoint broadcasting, a transient detector look-ahead can be employed on the encoder side. Having this additional information, time/frequency grids spanning across borders of granules can be comprised. This scheme provides a more flexible division of the time/frequency grids, and enables the system to work on a constant bitrate basis. Referring to Fig. 3, the granules are divided into eight subgranules. The transient detector operates on granules with the same timespan as the granule that overlap 50 % of two consecutive granules, that is, the transient detector look-ahead is half a granule. The transient detector has detected a transient in subgranule 6 at time n-1, and a transient in subgranule 7 at time n. With these values as indices into the table, the corresponding time/frequency grid for granule n might be as shown in Fig. 3c. As seen from the figure, subgranule 7 of the granule at time n-1 is included in the time/frequency grid of granule n.

Some examples of time segment grouping are given in Fig 4, where the subgranules are numbered from 000 to 111. L denotes low frequency resolution and H denotes high resolution. In the example the number of scalefactors in a high resolution segment is assumed to be two times that of a low resolution segment. If no transient is present in or next to a specific granule, the granule is divided into two high resolution segments of equal length and the scalefactors are calculated, Fig 4a. The scalefactor matrix relative size is shown in the figures, using this case as reference. The two control signals, *tran_flag* and *tran_pos*, are also shown in the figure. (Here one of the "extra states" mentioned above has been utilized.) If the two sets of scalefactors in Fig 4a do not differ more than a certain amount, only one set of high resolution scalefactors is sent, Fig 4b. Fig 4c – 4f show some cases where a transient, denoted by <T> is present. N and P indicate segments spanning over a granule border, where N means that the corresponding scalefactors are sent in next granule and P that they were sent in the previous granule. Notice that using this scheme, in many cases a transient does not generate more scalefactors than the reference case (Fig 4c, e, and f). Obviously, it is possible to design a scheme that keeps the matrix size constant, if desired. For a typical programme material, the transient indexing system has a performance similar to that of a system using a constant time update step of T_q and constant high frequency resolution, Fig 4g. If a high resolution segment corresponds to 20 scalefactors and the average number of bits per scalefactor again is taken as 3, the static system generates an average of 20*3*8 = 480 bits/granule (no signal bits required). Assuming that the state in Fig 4b uccurs 25% of the time in average and the Fig 4d class of states with relative data size 1.5 also occur 25% of the time, the bitrate of the dynamic system computes to (0.25*0.5 + 0.25*1.5 + 0.5*1)*20*3*2 + 4 = 124 bits/granule, i.e. an average of only 26% of that of the static system. Hence a major data reduction is achieved when using the dynamic grouping of scalefactors according to the invention.

Time/Frequency Switched Scalefactor Encoding

Fourier analysis states:

$$\mathfrak{F}[\delta(t)] = 1 \qquad \text{(Eq 9)}$$

$$\mathfrak{F}[1] = 2\pi\delta(\omega) \qquad \text{(Eq 10)}$$

This implies that a pulse in the time domain corresponds to a flat spectrum in the frequency domain, and a "pulse" in the frequency domain, i.e. a single sinusoidal, corresponds to a stationary signal in the time domain. In other words a signal is never transient in two domains simultaneously. In a spectrogram, i.e. a time/frequency matrix display, this property is evident, and can advantageously be used when coding spectral envelopes. A tonal stationary signal can have a very sparse spectrum not suitable for delta coding in the frequency-direction, but well suited for delta coding in the time-direction, and vice versa. This is displayed in Fig. 5. Throughout the following description a vector of scale factors calculated at time n_0 represents the spectral envelope

$$Y(k,n_0)=[a_1, a_2, a_3, \ldots, a_k, \ldots, a_N], \qquad \text{(Eq 11)}$$

where $a_1 \ldots a_N$ are the amplitude values for different frequencies. Common practice is to code the difference between adjacent values in the frequency-direction at a given time, which yields:

$$D(k,n_0)=[a_2-a_1, a_3-a_2, \ldots, a_N-a_{(N-1)}]. \qquad \text{(Eq 12)}$$

In order to be able to decode this, the start value a_1 needs to be transmitted. As stated above this delta-coding scheme can prove to be most inefficient if the spectrum only contains a few stationary tones. This can result in a

delta coding yielding a higher bit rate than regular PCM coding. In order to deal with this problem, a time/frequency switching method, hereinafter referred to as T/F-coding, is proposed: The scalefactors are quantized and coded both in the time- and frequency-direction. For both cases, the required number of bits is calculated for a given coding error, or the error is calculated for a given number of bits. Based upon this, the most beneficial coding direction is selected.

As an example, DPCM and Huffman redundancy coding can be used. Two vectors are calculated, D_f and D_t:

$$D_f(k,n_0)=[a_2-a_1, a_3-a_2, \ldots, a_N-a_{(N-1)}], \qquad \text{(Eq 13)}$$

$$D_t(k,n_0)=[a_1(n_0)-a_1(n_0-1), a_2(n_0)-a_2(n_0-1), \ldots, a_N(n_0)-a_N(n_0-1)] \qquad \text{(Eq 14)}$$

The corresponding Huffman tables, one for the frequency direction and one for the time direction, state the number of bits required in order to code the vectors. The coded vector requiring the least number of bits to code represents the preferable coding direction. The tables may initially be generated using some minimum distance as a time/frequency switching criterion.

Start values are transmitted whenever the spectral envelope is coded in the frequency direction but not when coded in the time direction since they are available at the decoder, through the previous envelope. The proposed algorithm also require extra information to be transmitted, namely a time/frequency flag indicating in which direction the spectral envelope was coded. The T/F algorithm can advantageously be used with several different coding schemes apart from DPCM and Huffman, such as ADPCM, LPC etc.

When coding the spectral envelope for SBR the circumstances are somewhat different compared to ordinary spectral envelope coding. The replicated signal in the decoder has a formant structure and envelope created by the transposer. The received envelope is to be used for adjustment of the replicated signal. This means that its possible to use redundancy between the source area and the high band, i.e. instead of delta coding of adjacent scale factors, scalefactors are delta coded on an octave basis.

Practical implementations

An example of the encoder side of the invention is shown in Fig. 6. The analogue input signal is fed to an A/D-converter 601, forming a digital signal. The digital audio signal is fed to a perceptual audio encoder 602, where source coding is performed. In addition, the digital signal is fed to a transient detector 603 and to an analysis filterbank 604, which splits the signal into its spectral equivalents (subband signals). The transient detector could operate on the subband signals from the analysis bank, but for generality purposes it is here assumed to operate on the digital time domain samples directly. The transient detector divides the signal into granules and determines, according to the invention, whether subgranules within the granules is to be flagged as transient. This information is sent to the envelope grouping block 605, which specifies the time/frequency grid to be used for the current granule. According to the grid, the block combines the uniform sampled subband signals, to form the non-uniform sampled envelope values. As an example, these values might be the average or maximum energy for the subband samples combined. The envelope values are, together with the grouping information, fed to the envelope encoder block 606. This block decides in which direction (time or frequency) to encode the envelope values. The resulting signals, the

output from the audio encoder, the wideband envelope information, and the control signals are fed to the multiplexer 607, forming a serial bitstream that is transmitted or stored.

The decoder side of the invention is shown in Fig. 7. The demultiplexer 701 restores the signals and feeds the appropriate part to an audio decoder 702, which produces a low band digital audio signal. The envelope information is fed from the demultiplexer to the envelope decoding block 703, which, by use of control data, determines in which direction the current envelope are coded and decodes the data. The low band signal from the audio decoder is routed to the transposition module 704, which generates a replicated high band signal consisting of one or several harmonics from the low band signal. The high band signal is fed to an analysis filterbank 706, which is of the same type as on the encoder side. The subband signals are combined in the scalefactor grouping unit 707. By use of control data from the demultiplexer, the same type of combination and time/frequency distribution of the subband samples is adopted as on the encoder side. The envelope information from the demultiplexer and the information from the scalefactor grouping unit is processed in the gain control module 708. The module computes gain factors to be applied to the subband samples before recombination in the synthesis filterbank block 709. The output from the synthesis filterbank is thus an envelope adjusted high band audio signal. This signal is added to the output from the delay unit 705, which is fed with the low band audio signal. The delay compensates for the processing time of the high band signal. Finally, the obtained digital wideband signal is converted to an analogue audio signal in the digital to analogue converter 710.

CLAIMS

1. A method for spectral envelope coding in a source coding system where said system comprises an encoder representing all operations performed prior to storage or transmission, and a decoder representing all operations performed after storage or transmission, **characterised by**:

at said encoder, perform a statistical analysis of the input signal,

based on the outcome of said analysis, select the instantaneous time and frequency resolution to be used in the spectral envelope representation,

using said resolution, generate scalefactors representing said spectral envelope,

transmit said scalefactors together with a control signal describing said resolution,

at said decoder, using said control signal and said scalefactors in the synthesis of the output signal.

2. A method according to claim 1, **characterised in** that said instantaneous time and frequency resolution is obtained by grouping of elements in a time/frequency representation of said input signal, and calculating a scalefactor for every one of said groups.

3. A method according to claim 2, **characterised in** that said time/frequency representation is generated by a filterbank.

4. A method according to claim 3, **characterised in** that said filterbank is of fixed size.

5. A method according to claim 1, **characterised in** that said analysis employs a transient detector.

6. A method according to claim 5, **characterised in** that said instantaneous resolution is switched from a default combination of higher frequency resolution and lower time resolution to a combination of lower frequency resolution and higher time resolution at the onset of a transient.

7. A method according to claim 1, **characterised in** that said control signal describes positions within a granule of constant update rate, generated by said analysis, and said instantaneous resolution is chosen based on the positions within current and neighbouring granules, by the use of rules available to both said encoder and said decoder.

8. A method according to claim 7, **characterised in** that at most one position per granule is signalled.

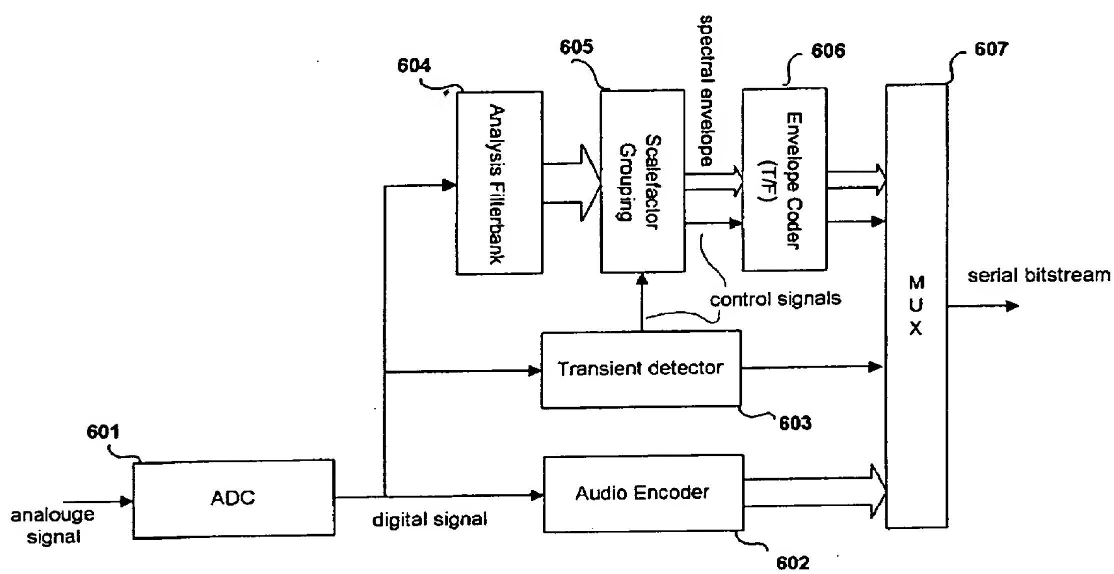
9. A method according to claim 1, **characterised in** that said scalefactors are coded both in the time and frequency direction, the momentarily most beneficial direction is determined, said most beneficial direction is used for said transmission.

10. A method according to claim 9, **characterised in** that the direction which generates the least coding error for a given number of bits is chosen.

11. A method according to claim 9, **characterised in** that the direction which generates the least number of bits for a given coding error is chosen.

12. A method according to claim **11**, **characterised in** that lossless coding is employed and separate tables are used for said time and frequency directions, in particular where said tables are used for selection of coding direction.

EFFICIENT SPECTRAL ENVELOPE CODING USING DYNAMIC SCALEFACTOR GROUPING AND TIME/FREQUENCY SWITCHING



ABSTRACT

The present invention provides a new method and an apparatus for spectral envelope encoding. The invention teaches how to perform and compactly signal a time/frequency mapping of the envelope representation, and further, encode the spectral envelope data efficiently using adaptive time/frequency direction coding. The method is applicable in both natural audio coding and speech coding systems and is especially suited for coders using SBR [WO 98/57436].

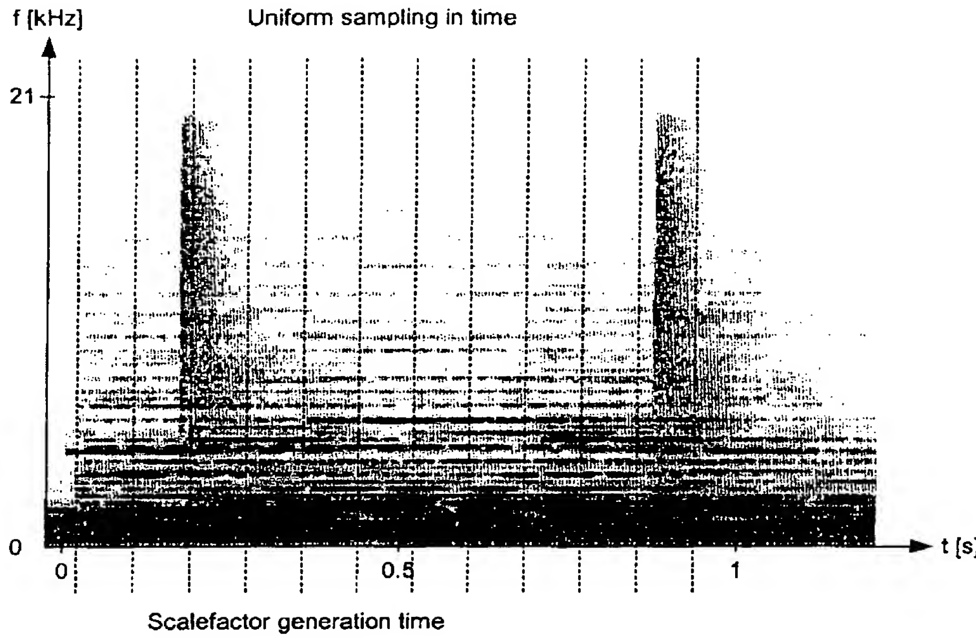

f [kHz]
Uniform sampling in time
21
0
0
0.5
1
t [s]
Scalefactor generation time

Fig. 1a

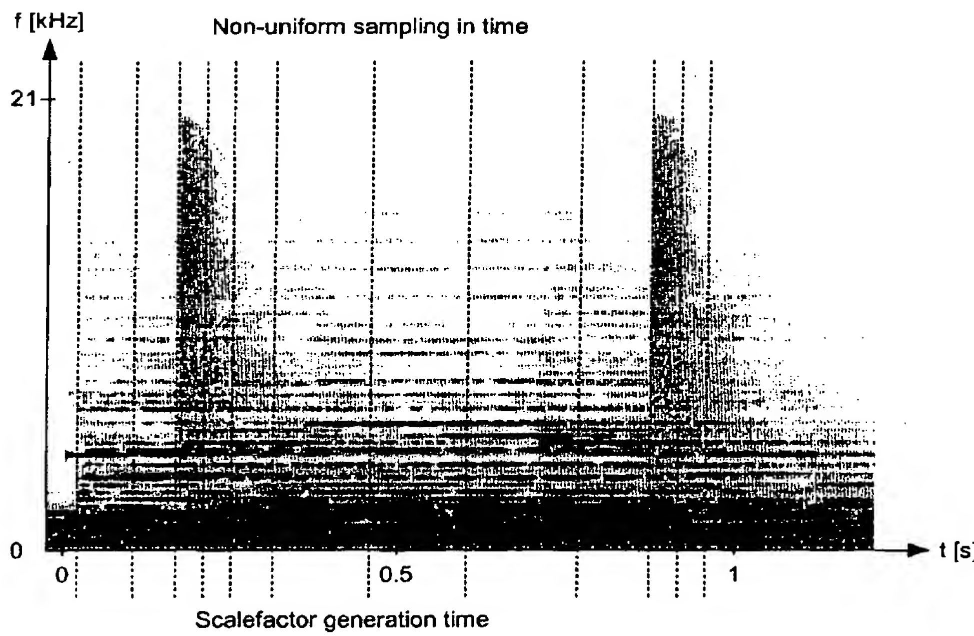

f [kHz]
Non-uniform sampling in time
21
0
0
0.5
1
t [s]
Scalefactor generation time

Fig. 1b

```
    +---------------+---------------+
    |0!1!2!3!4!5!6!7|0!1!2!3!4!5!6!7| ( 8 steps per granule)
    +---------------+---------------+
    |0!1!2!3!4!5!6!7!8!9!1!1!1!1!1!1| (16 steps per granule)
    | ! ! ! ! ! ! ! ! ! !0!1!2!3!4!5|
+---+---------------+---------------+
|  0|T! ! ! ! ! ! ! |T! ! ! ! ! ! ! |
|  1|T! ! ! ! ! ! ! | !T! ! ! ! ! ! |
  :  :                  . . . .
|  6|T! ! ! ! ! ! ! | ! ! ! ! ! !T! |
|  7|T! ! ! ! ! ! ! | ! ! ! ! ! ! !T|
|  8|T! ! ! ! ! ! ! | ! ! ! ! ! ! ! |
|  9| !T! ! ! ! ! ! | !T! ! ! ! ! ! |
  :    :                . . . . .
| 15| !T! ! ! ! ! ! | ! ! ! ! ! ! !T|
| 16| !T! ! ! ! ! ! | ! ! ! ! ! ! ! |
| 17| ! !T! ! ! ! ! | ! !T! ! ! ! ! |
  :      :                . . . .
| 22| ! !T! ! ! ! ! | ! ! ! ! ! ! !T|
| 23| ! !T! ! ! ! ! | ! ! ! ! ! ! ! |
| 24| ! ! !T! ! ! ! | ! ! !T! ! ! ! |
  :        :                . . .
| 28| ! ! !T! ! ! ! | ! ! ! ! ! ! !T|
| 29| ! ! !T! ! ! ! | ! ! ! ! ! ! ! |
| 30| ! ! ! !T! ! ! | ! ! ! !T! ! ! |
  :          :                . .
| 33| ! ! ! !T! ! ! | ! ! ! ! ! ! !T|
| 34| ! ! ! !T! ! ! | ! ! ! ! ! ! ! |
| 35| ! ! ! ! !T! ! | ! ! ! ! !T! ! |
| 36| ! ! ! ! !T! ! | ! ! ! ! ! !T! |
| 37| ! ! ! ! !T! ! | ! ! ! ! ! ! !T|
| 38| ! ! ! ! !T! ! | ! ! ! ! ! ! ! |
| 39| ! ! ! ! ! !T! | ! ! ! ! ! !T! |
| 40| ! ! ! ! ! !T! | ! ! ! ! ! ! !T|
| 41| ! ! ! ! ! !T! | ! ! ! ! ! ! ! |
| 42| ! ! ! ! ! ! !T| ! ! ! ! ! ! !T|
| 43| ! ! ! ! ! ! !T| ! ! ! ! ! ! ! |
| 44| ! ! ! ! ! ! ! |T! ! ! ! ! ! ! |
| 45| ! ! ! ! ! ! ! | !T! ! ! ! ! ! |
  :                       . . . .
| 50| ! ! ! ! ! ! ! | ! ! ! ! ! !T! |
| 51| ! ! ! ! ! ! ! | ! ! ! ! ! ! !T|
| 52| ! ! ! ! ! ! ! | ! ! ! ! ! ! ! |
+---+---------------+---------------+
```

Fig. 2

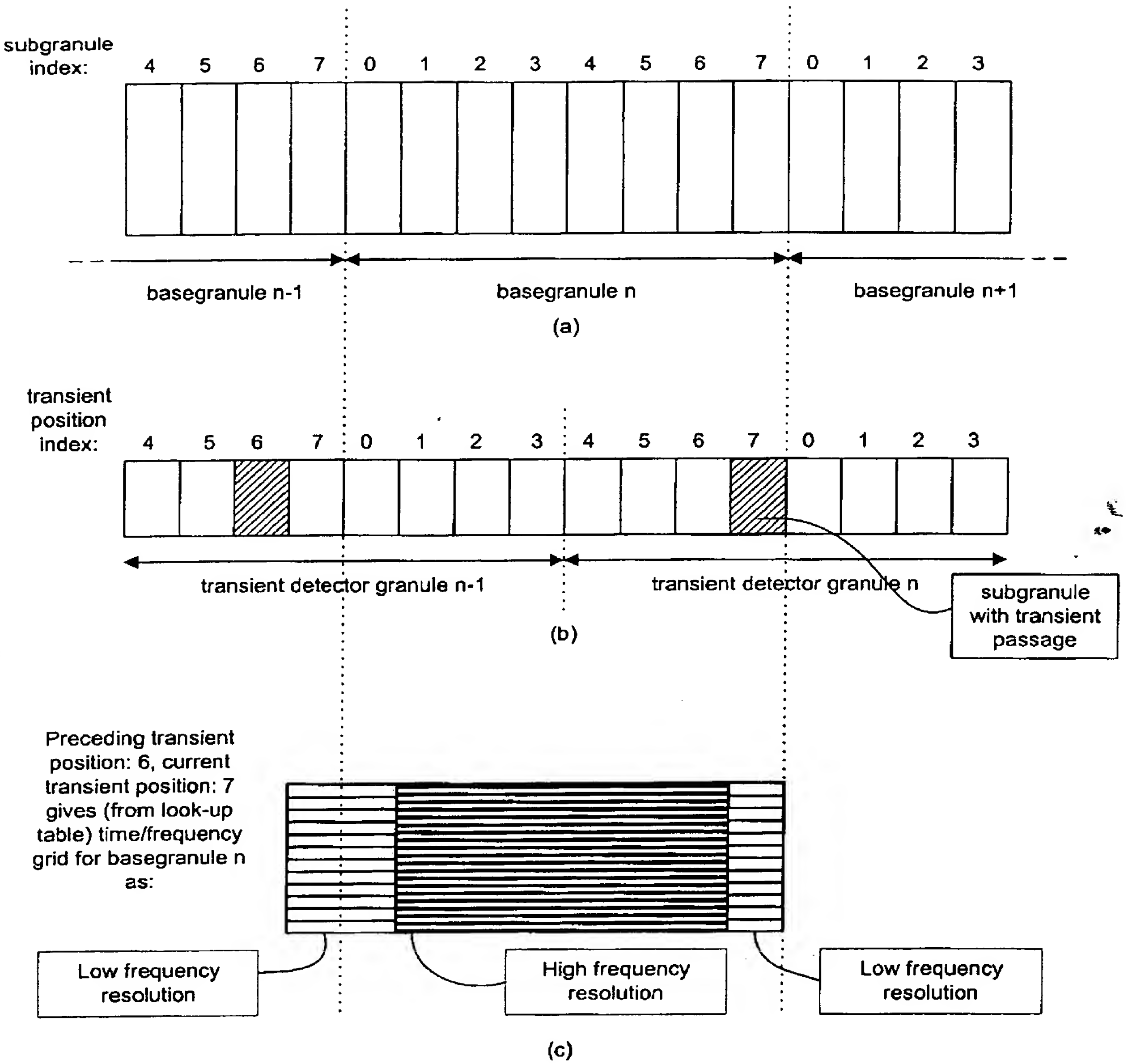

subgranule index:
4 5 6 7 0 1 2 3 4 5 6 7 0 1 2 3
basegranule n-1
basegranule n
basegranule n+1
(a)
transient position index:
4 5 6 7 0 1 2 3 4 5 6 7 0 1 2 3
transient detector granule n-1
transient detector granule n
subgranule with transient passage
(b)
Preceding transient position: 6, current transient position: 7 gives (from look-up table) time/frequency grid for basegranule n as:
Low frequency resolution
High frequency resolution
Low frequency resolution
(c)

Fig. 2

```
                          <T>
a)  |000|001|010|011|100|101|110|111|     tran_flag = 0, tran_pos = 100
    |HHHHHHHHHHHHHHH|HHHHHHHHHHHHHHH|     rel data size = 1.0 (reference)

b)  |000|001|010|011|100|101|110|111|     tran_flag = 0, tran_pos = 000
    |HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH|     rel data size = 0.5

                      <T>
c)  |000|001|010|011|100|101|110|111|     tran_flag = 1, tran_pos = 100
    |HHHHHHHHHHHHHHH|LLLLLLL|LLLLLLL|     rel data size = 1.0

              <T>
d)  |000|001|010|011|100|101|110|111|     tran_flag = 1, tran_pos = 010
    |HHHHHHH|LLLLLLL|LLLLLLL|HHHHHHH|     rel data size = 1.5

                  <T>
e)  |000|001|010|011|100|101|110|111|     tran_flag = 1, tran_pos = 011
    |HHHHHHHHHHH|LLLLLLL|LLLLLLL|NNN|     rel data size = 1.0

          <T>
f)  |000|001|010|011|100|101|110|111|     tran_flag = 1, tran_pos = 001
    |PPP|LLLLLLL|LLLLLLL|HHHHHHHHHHH|     rel data size = 1.0

g)  |000|001|010|011|100|101|110|111|     no control signals
    |HHH|HHH|HHH|HHH|HHH|HHH|HHH|HHH|     rel data size = 4.0
```

Fig. 4

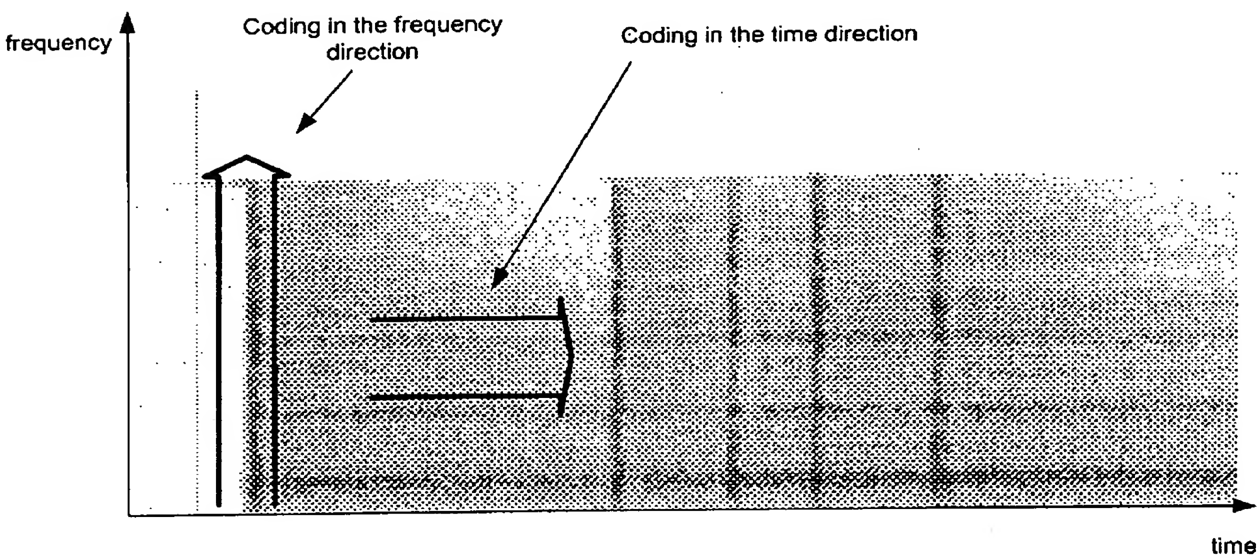
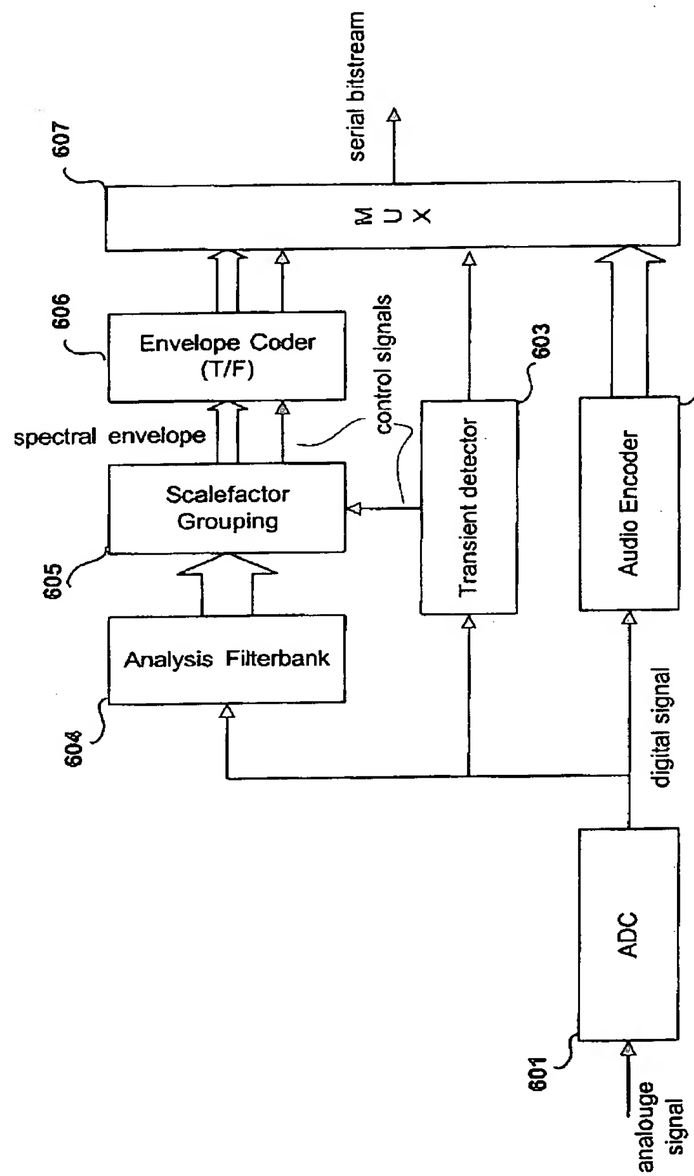


Fig. 5



serial bitstream
607
M U X
606
Envelope Coder (T/F)
control signals
spectral envelope
Scalefactor Grouping
605
Analysis Filterbank
604
Transient detector
603
Audio Encoder
602
digital signal
ADC
601
analouge signal

Fig. 6

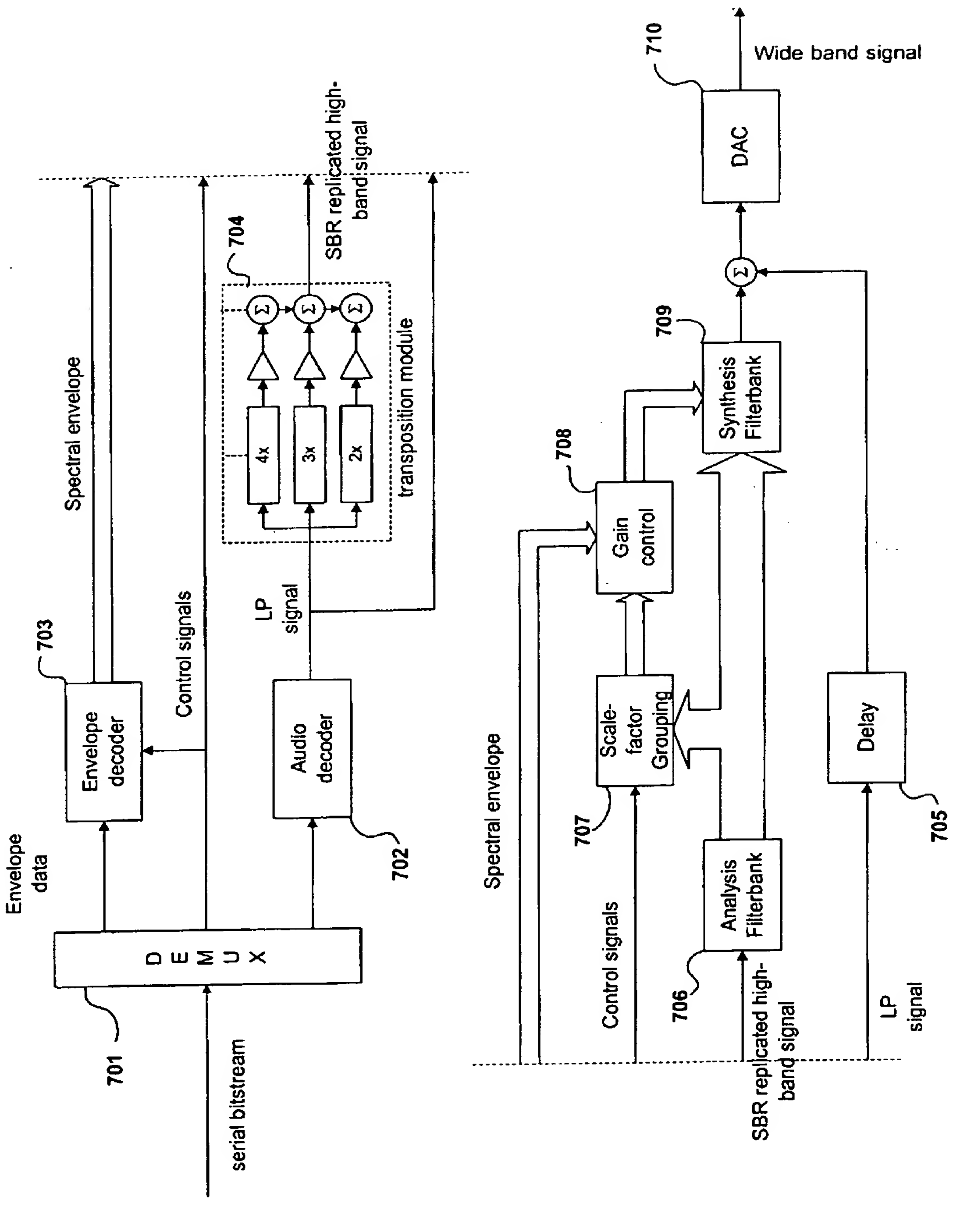

serial bitstream
701
DEMUX
Envelope data
703
Envelope decoder
Spectral envelope
Control signals
Audio decoder
702
LP signal
704
4x
3x
2x
Σ
transposition module
SBR replicated high-band signal
Spectral envelope
Control signals
SBR replicated high-band signal
LP signal
706
Analysis Filterbank
707
Scale-factor Grouping
708
Gain control
709
Synthesis Filterbank
705
Delay
Σ
710
DAC
Wide band signal

Fig. 7

THIS PAGE BLANK (USPTO)